Interaction Guidelines for Personal Voice Assistants in Smart Homes

Tamino Huxohl
Applied Informatics Group
CITEC - Bielefeld University
Bielefeld, Germany
thuxohl@techfak.uni-bielefeld.de

Marian Pohling
Ambient Intelligence Group
CITEC - Bielefeld University
Bielefeld, Germany
mpohling@techfak.uni-bielefeld.de

Birte Carlmeyer
Applied Informatics Group
Dialogue System Group
CITEC - Bielefeld University
Bielefeld, Germany
bcarlmey@techfak.uni-bielefeld.de

Britta Wrede
Applied Informatics Group
CITEC - Bielefeld University
Bielefeld, Germany
bwrede@techfak.uni-bielefeld.de

Thomas Hermann
Ambient Intelligence Group
CITEC - Bielefeld University
Bielefeld, Germany
thermann@techfak.uni-bielefeld.de

***Abstract*—The use of Personal Voice Assistants (PVAs)[1] such as Alexa and the Google Assistant is rising steadily, but there is a lack of research investigating common issues and requests of PVAs in the context of smart home control. We address this research question with an online survey ($n = 65$), using a qualitative evaluation of users' problems and improvement requests. Our analysis leads to a partly hierarchic clustering of issues & recommendations for interaction capabilities of PVAs into seven basic categories, allowing us in turn to derive implications and to condense them into design guidelines for future Human-Agent Interaction (HAI) with PVAs. Specifically, we formulate and elaborate the concepts *Authentication & Authorization*, *Activity-Based Interaction*, *Situated Dialogue*, and *Explainability & Transparency* as key topics for making progress towards smooth interaction with smart homes.**

***Index Terms*—personal voice assistants, smart home, interaction guidelines**

I. Introduction and Related Work

In the last few years, cloud computing made the power of deep neural networks and big data accessible via low budget end-user devices [2]. As a result, according to the market research firm Tractica, 1.3 billion PVAs such as Alexa, the Google Assistant and Siri already moved into users' households and the amount is estimated to increase to 1.8 billion by 2021 [3]. Speech recognition has reached a quality that is sufficient to get accepted by end users and speech interaction has become a common interaction modality in many users' daily life [4], [5]. A huge potential has been identified in the area of Internet of Things (IoT) and smart homes where PVAs are used to control smart appliances [6] or even autonomously take control of the environment [7]. Recently, Microsoft performed a user study with 2000 participants evaluating the current state and future perspective of PVAs [8]. They focused on usage trends as well as data security concerns and discovered that 54% of users manage their home with a smart speaker.

However, the literature reveals a lack of research in the evaluation of interaction capabilities of PVAs to configure and control smart homes. Studies investigating common issues with using PVAs were only published recently. For example, Budiu and Laubheimer evaluated the usability of PVAs in a user study with 17 participants [9]. Each participant performed a set of tasks using Alexa, the Google Assistant or Siri and was interviewed afterwards. Of special interest were the six User Interface (UI) techniques: voice input, natural language, voice output, intelligent interpretation, agency, and integration. Except in the category *voice input*, the PVAs performed poorly in all UI techniques ('bad' or 'terrible') [9]. López et al. evaluated the performance of Siri, Alexa, Cortana and the Google Assistant [10]. Eight participants rated the naturalness and the correctness of the assistant's answers to questions targeting different service areas such as music control, navigation and information requests. In this study, interactions with the Google Assistant were rated as the most natural. Both user studies provide interesting insights into usability issues. However, a broader breakdown of current issues with PVAs in the context of smart home control is missing. Yet, it might yield important insights regarding further research demands.

Purington et al. investigated reviews of Alexa posted on Amazon and drew conclusions on how users perceive and interact with Alexa as a social and conversational agent. They found a link between the personification of Alexa and the user's satisfaction [11]. Furthermore, users who mentioned technical issues were less satisfied with the device. Similarly, Williams et al. investigated the perceived satisfaction of Alexa users in an online survey [12]. Like Purington et al., they do not go into detail about the mentioned issues and problems. Both studies only focus on Alexa and do not incorporate other

[1]For an overview of classification and terminology see [1].

978-1-7281-0984-8/19/$31.00 ©2019 IEEE

PVAs.

Besides these investigations of PVAs, little research exists that addresses design implications and provides design guidelines for smart home control. Caivano et al. present an extensive literature review about various smart home control support [13], on the basis of which they developed a set of design implications for future smart home control devices.

Lee et al. designed a wizard of Oz study concept to develop early state design guidelines for PVAs [14]. In a test study with 8 participants, they were able to observe the influence of personality and emotions of PVAs on the interaction success. However, they mainly focused on the developed study concept to prototype design guidelines rather than researching new guidelines because the sample size of their study was not large enough to generalize their results.

Coskun et al. carried out semi-structured interviews with 20 participants in order to investigate user expectations and benefits of various features for smart household appliances [15]. Based on these insights, they also present several design principles. Indeed, their research provides fundamental principles for smart home development, but interestingly they do not take PVAs as a UI into account.

In summary, the literature shows a lack of research investigating current issues with—and improvement requests for—PVAs in the context of smart home control. Yet, it could unveil fundamental principles for smart home development. Hence, we address the following research question in this paper: Which issues are the most common when using PVAs for smart home control? For that, we carried out an online survey to identify the most annoying problems for smart home inhabitants. The online survey allowed us to reach participants who were experienced with various PVAs. Their answers provide an overview of current issues in the HAI from the users' perspective. We analyzed these issues and improvement requests and categorized them. Based on the gained insights, we propose some resulting design guidelines for further development of PVAs.

II. Method

The data analyzed in this paper is part of a study whose main objective was to find a set of capabilities of smart homes that users want to access via their PVA. The study was conducted as an online survey. It contained a broad array of questions which allows further analysis regarding the research question of this paper. The following section describes the survey and the subsequent data analysis.

A. Structure of the Online Questionnaire

The questionnaire consisted of five parts, the last of which being a repetition of part three:

1) *Demographics*
2) *Experience*
3) *Problems and Improvements*
4) *Author's Suggestions*
5) *Problems and Improvements*

The first part *Demographics* contained questions about the participant's gender, age, country of residence and occupation. The second part *Experience* focused on the participant's experience with PVAs and smart homes. For example, participants were asked to rate how experienced they were with Alexa, the Google Assistant, Siri and Cortana. Additionally, they could provide information on which capabilities of PVAs they use and how many smart speakers they own. Regarding experience with smart homes, participants rated how familiar they are with smart homes, how well they can imagine living in a fully connected home and reported their approval of an assistant independently controlling smart home appliances.

In the third and fifth part *Problems and Improvements*, participants could note up to two problems with—and improvements for—the usage of PVAs in smart homes. We discriminate between problems and improvements as follows: Problems occur if a functionality does not work as expected while improvements can be the addition of a new feature. It is important to note, however, that there is not a hard boundary between problems and improvements. This is because the solution to a problem can also be seen as an improvement. For example, errors in the speech recognition are a problem while improving the speech recognition to produce fewer errors is certainly an improvement.

The *Problems and Improvements* parts were repeated in step 3 and 5 because of part four, *Author's Suggestions*. This part consisted of problems and improvements suggested by the author of the study. Hence, the *Problems and Improvements* section in part 3 allows to elicit an unbiased response while the repeated section in part 5 could reveal other kinds of problems and improvements because of reading about the suggestions of the author in part 4. However, a bias could not be observed since most of the problems (96 out of 115) and most of the improvements (74 out of 94) were mentioned in the third part. Thus, and because part four was added for the original objective of the study, the suggestions of the author are not detailed in this paper. At the end of the survey, participants had the opportunity to leave comments.

B. Procedure

The survey was online from 2018-06-05 to 2018-07-01. It was available in German, since it is the first language of the authors, and in English to reach an international audience. Because participants were asked about problems with—and improvements for—the usage of PVAs in smart homes, it was a prerequisite that they had some experience with this topic. Thus, the survey was published in the following ways:

- Friends and acquaintances who own a smart speaker were asked to complete the survey.
- It was published via a mailing list of the Cognitive Service Robotics Apartment (CSRA) research project [16] with the request to forward it to people with interest in the topic.
- It was posted in the following online forums:
  - www.reddit.com/r/homeautomation
  - www.reddit.com/r/smarthome

- www.reddit.com/r/googlehome
- www.gassistant.de/forum
- www.smarthomeforum.de

As an incentive, a Google Home Mini was raffled among participants completing the questionnaire.

C. Data Analysis

Altogether, 65 participants completed the survey by submitting their responses. However, the survey contained a control question in the form of asking twice how many smart speakers a participant owns. Four of the 65 participants provided a different answer the second time the question was asked. Hence, their responses were excluded from the evaluation.

To analyze problems with—and improvements for—PVAs in the context of smart home control, we carried out a thematic analysis. Therefore, three experts in this field individually generated an initial coding scheme to categorize the problems and improvements noted by the participants. Subsequently, these experts jointly identified the applicable categories. If applicable, categories were further divided into subcategories. The resulting coding scheme was again verified by a fourth expert.

Regarding the categorization it is important to note that some problems and improvements did not match a distinct category. Therefore, they were put into multiple categories which means that the numbers do not exactly add up to the totals. For instance, one participant noted the problem "Tedious start phrases and the inability to change those phrases. It is sometimes difficult to have it hear you properly the first time". This statement entails two problems: On the one hand, the PVA is missing the option to customize the hotword. On the other hand, the hotword is sometimes not recognized. So, this problem was put into two different categories. Note however, that this is not a common issue. Only three problems and four improvements were affected by this.

In addition to these ambiguous problems and improvements, some responses did not match any category, were unrelated to the topic or not a problem or improvement at all. For example, the improvement "lower cost" did not fit any of the categories and "nothing new, same as before" is not an improvement at all. We excluded 18 problems and 13 improvements for these reasons, leaving 97 problems and 81 improvements for the evaluation.

III. Results

In this section, the results of the online survey are presented. First, *demographic information* about the participants is given. This is followed up by an overview of the results of the *experience* section in which it is detailed how many smart speakers participants owned, how experienced they were with different PVAs as well as how how experienced they were with smart homes in general. The results indicate that the participants were experienced with the usage of PVAs for smart home control. Afterwards, the categorization of *issues and improvements requests* mentioned by the participants is presented.

A. Demographic Information

On average a participant was $\mu = 30.4$ years old with a minimum age of 19 years, a maximum age of 53 years and a standard deviation of $\sigma = 7.8$ years. Participants were mostly male (56) with only three females and two participants giving no response or identifying neither as male nor as female. With a count of 21, most participants worked in the IT sector. Eight participants were students, four were scientists and another four were unemployed. The remaining 24 participants were occupied in many different sectors. At the time of filling out the questionnaire, 39 participants resided in the United States of America, eleven in Germany and another eleven in countries that were not mentioned more than twice, which were other countries in Europe or New Zealand. Thus, the sample analyzed in this paper consists mostly of young male adults residing in the USA or in Europe.

B. Previous Experience

Figure 1 shows how many smart speakers each participant owned. Eight participants did not own a smart speaker which means that 53 out of the 61 participants owned a smart speaker. For the participants of this study, it was common to have more than one smart speaker since 75.47% of participants who owned a smart speaker had more than one. A single participant even owned 16 smart speakers. On average a participant owned $\mu = 2.98$ smart speakers with a standard deviation of $\sigma = 2.72$.

As how experienced participants rated themselves with different PVAs is depicted in Figure 2a. The ratings were done on a scale from 1 (not experienced at all) to 5 (extremely experienced). All in all, participants were the most experienced with the Google Assistant. It received an average rating of $\mu = 3.75$ which conforms to being very experienced with it. Alexa received the second highest rating with an average of $\mu = 2.84$. Participants were less experienced with Siri ($\mu = 2.15$) and the least experienced with Cortana ($\mu = 1.85$).

In addition to rating their experience with different PVAs, participants also stated which functions of a PVA they use. The most utilized functions were *Playing Music*, *Setting Alarms or Timers* and *Requesting Information*. Each of these was used by over 50 participants. *Controlling Smart Home Appliances* such as light and thermostats were also frequently used functions. 46 participants stated that they control lights and 25 stated that they control thermostats via a PVA.

As described above, participants rated their experience with smart homes in three different questions. First, they rated their familiarity with smart homes, then how well they could imagine living in a fully connected home and at last if they approve of an independently acting assistant in a smart home. The results are depicted in Figure 2b. On average participants stated to be very familiar ($\mu = 3.64$) with smart homes. Furthermore, they could imagine living in a fully connected home with an average rating of $\mu = 4.36$ and they approve of an independently acting assistant with an average rating of $\mu = 3.69$.

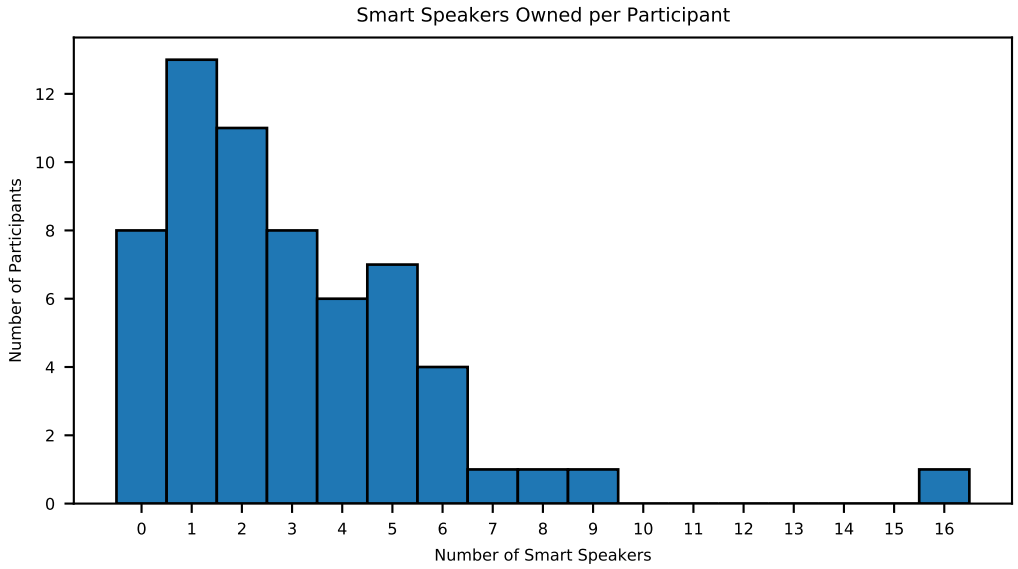

Fig. 1. Histogram displaying how many participants owned how many smart speakers.

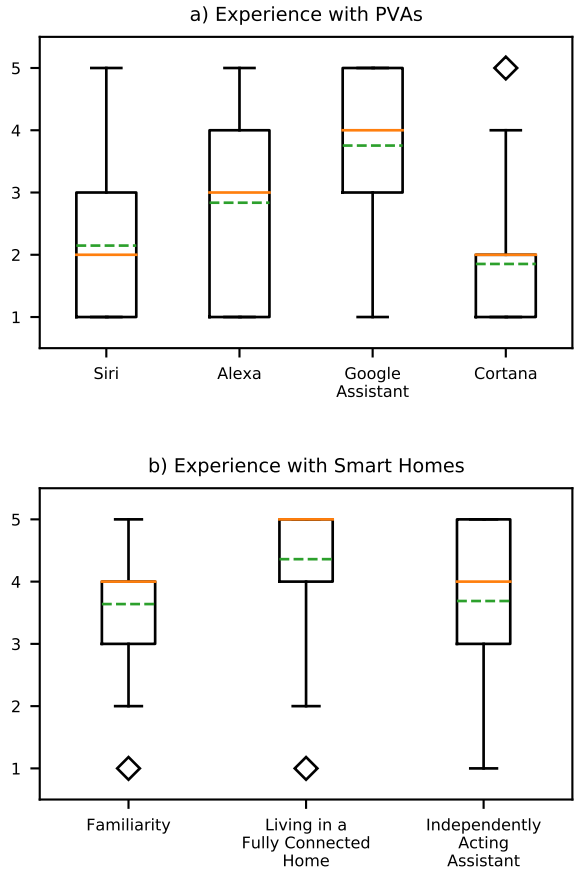

Fig. 2. Boxplots displaying as how experienced participants rated themselves with different PVAs (a) and smart homes (b). The solid line is the median and the dotted line is the average. The diamonds represent outliers which means that they are outside of [5,95] percentiles.

To summarize, participants were experienced with PVAs in the context of smart home control. They rated themselves to be very experienced with the Google Assistant and most of them used a PVA to control smart home appliances. Besides, they were experienced with smart homes and approved of an independently acting assistant. Therefore, participants were part of the desired focus group and thus qualified to provide issues with—and suggest improvements for—the usage of PVAs in smart homes from a user perspective.

C. Issues and Improvement Requests

For further analysis, we categorized the collected problems and improvement requests noted by the participants as follows:

- Language Processing
  - Command Execution
  - Dialogue Context
  - Hotword
  - General
- Capabilities
  - Hardware
  - Software
- User Context
  - Adaptability
  - Customization
- Feedback
- Capability Communication
- Privacy & Security
- Connectivity

Figure 3 depicts the number of problems & improvements noted by category. The category *Capabilities* contains capabilities that participants want to have but that are not yet available or need improvement. These capabilities are either part of the

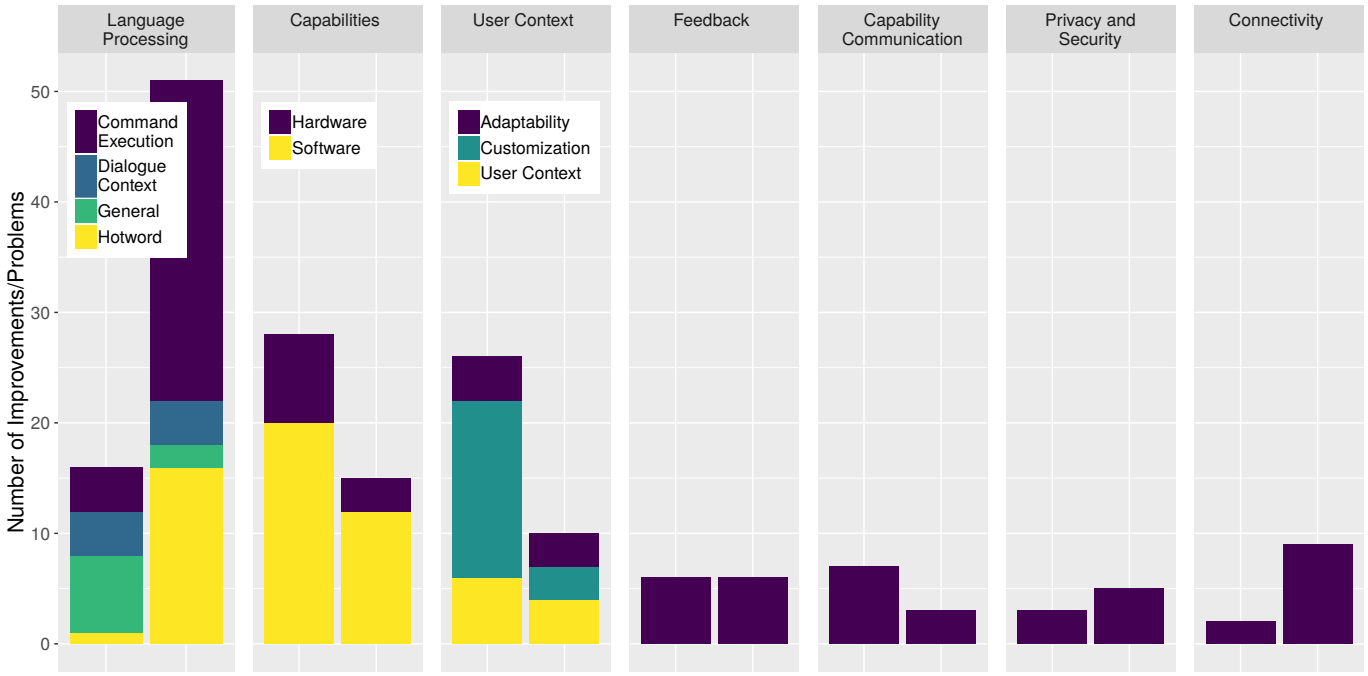

Fig. 3. Bar chart illustrating the categorization of problems and improvement requests. Bars are grouped by category and grey tones indicate the size of subcategories. For each category, the bar on the left indicates the amount of improvements and the bar on the right the amount of problems in it.

Hardware subcategory ("lack of native support for some devices", "better microphones") or *Software* subcategory ("can't send texts through the Google Home", "add an intercom or walkie talkie ability"). This category contains 15 problems and 28 improvements, making it the largest category regarding improvements. Most of the problems and improvements fall into the *Software* subcategory.

The category *Capability Communication* contains three problems and seven improvement requests of users who are unsure about the capabilities of their PVA. They do not know which commands are supposed to work ("I do not have a good overview of the capabilities of the smart speaker. And I never know if I didn't use the right words or if a functionality is not supported."). Consequently, some participants wish to have some kind of documentation ("a big registry with a lot of voice commands").

The category *Connectivity* comprises two kinds of issues and improvements: On the one hand, smart home devices lose connection and therefore, do not respond to commands ("Device does not turn on/off because of connectivity"). On the other hand, some participants want to have a PVA that does not require a connection to the Internet ("A "local-only" mode"). Overall, this category includes nine problems and two improvements.

The category *Feedback* contains six problems as well as six improvements. They are about auditory system feedback being too loud, too quiet, unnecessary altogether or even semantically wrong ("Report item is offline after turning the item off or on"). In addition, a few participants want to be able to customize system feedback ("Better/customization feedback loop, e.g. I do not want feedback for all requests.").

The largest category regarding problems—52 out of 97—is *Language Processing*. It is also a sizable category regarding improvements containing 16. In general, it is about cases of flawed speech recognition of PVAs. Most of the problems and improvements in this category can be placed into one of three subcategories: The subcategory *Command Execution* is about commands being executed incorrectly or not all all ("Incorrect interpretation of commands."). The subcategory *Hotword* refers to the hotword either not being recognized (false negative) ("Trigger phrase not being heard") or being recognized even if the PVA was not addressed by the user (false positive) ("Accidentally thinks you were talking to it"). Finally, the subcategory *Dialogue Context* is about the interaction being unnatural. For example, PVAs do not use information from the foregone dialogue ("The fact that Google assistant can't keep a conversation running with you"). The remaining problems and improvements are more *General*, such as the discrimination of some groups of persons ("difficulty understanding kid voices", "adding more languages like polish for example") or general demands for improvement ("Improve voice understanding").

Privacy & Security is the least frequent category. It includes only five problems and three improvements. Privacy is a concern, since the PVA uploads audio to a non-local storage ("Privacy! These devices are recording your audio in your home and uploading it to non-local "cloud" storage."). Security is a concern, because unauthorized users could execute undesirable commands ("Offer the possibilities to lock functionalities to specific users (do not allow visitors to buy

goods, for example)").

Finally, the category *User Context* comprises ten issues and 26 improvements. Several of them directly concern the PVA not utilizing context information. For instance, the PVA should be able to recognize the speaker and act accordingly ("Learn who is speaking and tailor responses for them.") as well as to integrate the user's location into its actions ("Location-aware contexts: If I say turn off the lights it should know where I am in the house and turn off the lights in that room only."). Other issues and improvement requests belong to one of two subcategories: The subcategory *Adaptability* describes issues with the PVA not adapting its behaviour to the user, e.g. by learning rules from habits ("Automating actions that I regularly ask it to do such as turning off lights at a certain time."). The second subcategory *Customization* contains requests to customize the interaction with the PVA. The requests range from creating custom commands and individualizing the degree of feedback, to setting a custom hotword and changing the voice of the assistant ("More user flexibility (allow wide range of custom tasks, change trigger word, response phrases, etc)").

IV. Resulting Interaction Concepts and Guidelines

To deal with several of the issues and requests for improvements described in the previous section, we propose the following four interaction concepts/guidelines: *Authentication & Authorization*, *Activity-Based Interaction*, *Situated Dialogue*, and *Explainability & Transparency*. Even though we explicitly asked for issues and requests in the context of smart home control, several problems and resulting concepts emerged which are about interaction with PVAs or even HAI in general. In the following, each of them is described in detail and it is explained how and which issues and requests they solve.

A. Authentication & Authorization

As the title suggests the purpose of this guideline is twofold: First, a PVA should be able to identify the person it is interacting with, which is called authentication. Second, a PVA should verify the permissions of the user who issues a command, which is called authorization. Authorization requires authentication because permissions of a user can only be meaningfully verified if the user is identified beforehand. As an example for the advantages of this guideline, imagine a guest asking for today's appointments. Without authentication and authorization the PVA will inform the guest about the owner's appointments even though they may contain sensitive information such as doctor appointments. With it, however, the request could be denied because the user is either not identified or—as a guest—not authorized to access personal calendar information.

Implementing authentication is not an easy task because in a dialogue it has to be done solely based on the voice of the user. Despite this, most PVAs provide a voice recognition feature. In the case of Alexa it is called *Voice Profiles*[2] and in the case of Google it is called *Voice Match*[3]. For Siri the situation is different because it is not developed for a multi-user use case. Still, Siri attempts to recognize the voice of the owner of the device to prevent activation via voices from others [17]. Certainly, these voice recognition features are currently not secure. For instance, the documentation of the Google Assistant states that "a voice that sounds like yours might be able to get these results, too"[4] (results refers to personalized information such as the calendar of the speaker). An alternative to recognizing the speaker's voice is to query a passphrase. However, this lowers the usability because it adds an additional dialogue turn and is also not secure if the user has to say the passphrase out loud in the presence of other people. In contrast, the implementation of authorization can be achieved more easily. If the user is authenticated correctly, it only needs to be checked if the user has the required permissions to execute the command. This can be further improved by also adapting the response to the user.

Authentication & Authorization does not directly solve many of the issues and requests stated by the participants in the online survey. Only the problems and improvements about security in the category *Privacy & Security* are dealt with directly. For instance, it should prevent strangers from accessing sensitive data and it should prevent malicious audio from triggering commands. However, this guideline is also indirectly needed to solve issues and requests from other categories. Mainly, everything from the category *User Context* requires authentication, because features such as adaptability, customization and tailoring responses to the user depend on speaker identification.

In addition to our findings, there also exists other literature about security concerns which can be solved by *Authentication & Authorization*. For instance, Portet et al. investigated the acceptance of voice control of smart homes by the elderly [18]. Their results show that the acceptance suffers from the elderly fearing that anyone with access to the voice interface, including intruders, could control their home. This would not be possible if the user was identified before being allowed to execute commands. Besides, Hoy mentions several security and privacy issues [19]:

- Anyone with access to the device can gather personal information and ask it to perform tasks.
- Siri unlocked the door for any user standing outside and requesting the door to be opened [20].
- Anyone with access to Alexa was able to order items with the owner's Amazon account, such as a six-year-old that ordered a dollhouse [21].

[2]https://www.amazon.com/gp/help/customer/display.html?nodeId=202199440, retrieved 2019-04-24

[3]https://support.google.com/googlehome/answer/7342711?hl=en, retrieved 2019-04-24

[4]https://support.google.com/assistant/answer/9071681?visit_id=636916937603568225-133746324&rd=2, retrieved 2019-04-24

B. Activity-Based Interaction

The survey showed that PVAs lack context awareness and adaptability. For users the most important context is given by their *activity*, e.g. reading, sleeping, preparing a meal, etc., as it sufficiently represents the user context regarding smart home control and the terminology is already familiar to humans. Thus, we claim that anchoring user-system interaction on the user's activity will improve HAI. Nearly every action humans perform in their home can be described as an activity. Activities can be used as a simplified and action-related model of the user's current context while each activity can function as an anchor point to link related environmental requirements. Already in 2013, Nguyen et al. were convinced *"that, in order to make buildings truly adaptable and maximize efficiency and comfort, they need to be more aware to the activities of the users"* [22]. They discovered that user activities and behaviors are the most important input for building automation systems related to economy and energy saving. Based on these facts and our study results, we introduce an *activity-based domotic control concept* to improve naturalness, efficiency, customization and adaptability of interactions.

a) Naturalness: In our study, participants requested "*more natural trigger*" and "*more intuitive control*". Activity-based control offers the potential to increase the naturalness of HAI, since users are already familiar with the concept of activities and naming them in conversations is common. As a consequence, commands get more natural. This can be clarified through a comparison with commands in classical scene control. In scene control, commands such as "Activate scene: cooking", which directly refer to the technical construct *scene*, are common. This would probably never be said in Human-Human Interaction (HHI). In contrast, the according activity-based statement "I will prepare a meal now." is more natural and normal in HHI.

b) Efficiency: The main idea of activity-based control is to make the interaction more efficient by reducing the number of required interaction cycles to adjust the environment related to certain user activities. The activity itself establishes the common ground between user and PVAs. Especially smart home control can be improved where inhabitants control the environment just by referring to their currently performed activity. Users can pre-configure their individual set of actions to be executed by the PVAs when an activity is indicated. Therefore, activity-based control is comparable to the classical *smart home scene control* with a special focus on personalized user activities. Particularly the individual configuration offers advantages regarding user customization.

c) Customization: The participants requested customization capabilities of PVAs. With an activity-based interaction concept, inhabitants could personalize the actions to be executed by PVAs. In addition, users may want to customize actions according to the activity's target location, because the user's needs as well as the availability of smart services usually vary between different locations. The activity "working", for example, could switch on the desk lamp in the office, while the same activity triggered in the living room, turns off the TV while increasing the brightness of the ceiling lamp. Based upon the participants' request of more user context-aware PVAs, we recommend providing PVAs with automated user and target location identification capabilities. As a suggestion, user authentication could be used to identify the inhabitant, in order to automatically load their personalized action presets. Furthermore, our survey discovered that 75% of the owners of smart speakers own more than one, to cover different rooms. Thus, the PVAs placement itself can be used as target location for action preset selection—in case no location has been explicitly mentioned in the conversation. Therefore, activity-based interaction would enable PVAs to support context-oriented customization: once a user refers to an activity during an interaction, the PVA would be able to execute a personalized and location-specific set of actions.

d) Adaptability: Last but not least, participants requested more adaptation capabilities of PVAs so that the system can derive rules from the inhabitants habits and proactively support them. Since PVAs are limited regarding their sensing capabilities, it would probably be required to include external hardware components to attain such a complex goal. Anyway, we propose activities as a suitable starting point to cluster user behavior. Such clustering can in turn be instrumental to stimulate learning of activity/action sets. For instance, training samples could be generated by observing the ongoing interactions of smart appliances (i.e., the wished action set) right after inhabitants indicate to perform a certain activity. Furthermore, sensor information from smartphones and smart watches can be used to learn and predict activities in advance. For example, PVAs could learn to start whispering as soon as the smart watch is placed next to the beds charging station.

C. Situated Dialogue

The language processing provides plenty opportunities for improvement. As described in Section III-C, the majority of issues fall into the category *Language Processing*. Even though the speech recognition itself greatly improved in the last years, natural interaction is still an unsolved issue. In order to have more natural interactions with PVAs, the natural language processing needs to be enhanced. Therefore, we propose the integration of context information in order to facilitate situated dialogues, which is also in line with the *activity-based interaction* concept. This context information can be manifold.

a) Dialogue Context: An example here would be the use of the dialogue history. This would allow the possibility of longer conversations and follow-up questions. This has a two-fold benefit: another mention of the hotword is not necessary and it is possible to refer back to the previous dialogue act. The latter allow lots of new interaction styles, such as follow-up questions. Furthermore, more natural corrections and specifications of the last dialogue act become possible, e.g. "light on in the living room ... less bright, please".

b) User Context: Another example—a more fundamental change of interaction concepts with PVAs—would be the use of other modalities to incorporate the user context. As

well known in the human-agent and human-robot interaction community, dialogue is multi-modal. The integration of further sensors such as cameras to be aware of the human's attention is necessary. This additional information provides fundamental benefits for more natural interactions in various ways.

(1) The incorporation of pointing gestures or human eye-gaze can significantly improve reference resolution performance in language processing [23], which would allow the use of pronouns, such as "switch it on".

Furthermore, (2) information of the human's attention can be used for better addressee recognition [24]. Several mentioned issues are related to the hotword detection, which can be improved with a more natural addressee recognition.

Additionally, (3) this additional information allows the monitoring of the interaction partner. In order to have a meaningful interaction, a good deal of different feedback information should be taken into account [25]—whether to verify that the interaction partner understood the current dialogue act or to be able to react on a distracted interaction partner [26]. Feedback signals, such as gazing behaviour, head nods, and verbal back-channels allow the system further insight into the mental states of the user.

As already mentioned in the last part, (4) user activity is one aspect of context information. Depending on the current user activity, the system could react in different ways, e.g., speaking louder when the user takes a shower. Current systems already did a first step in this direction. The Google Assistant for example decrease the volume of the currently played back music, whenever the hotword is detected, which is beneficial for the speech recognition. The integration of the user activity provides more benefits, not only for the speech recognition but also for further processing modules. Based on the context, the same phrase could be interpreted differently. Furthermore, the resulting action could vary for the different activities.

Even though, the system would benefit from more context, we have to be aware that collecting more data—be it through additional sensors such as cameras or just gather more information from the speech—carries data privacy risks. Not every user wants e.g. a camera in his private area. A balance must be found between the possibility of natural interaction and data privacy.

D. Explainability & Transparency

Many users asked for more information about the system's capabilities and commands they can use. This addresses the more general issues explainability and transparency. This can be achieved by approaches that are either based on the system's perspective or on the user's perspective.

a) System's Perspective: As our sample represents a technically highly versatile subsample of the whole population most of the suggestions are based on existing explaining and transparency concepts which are based on the system's perspective. This means that the system provides information about existing processes, functionalities and commands that it knows. This entails two kinds of transparency: (1) on the one hand a post hoc transparency which requires explanations when a command is being processed or has failed. For example, it is practically impossible for a user to understand why a certain command could not be executed: because a wrong word was recognized, because it contained a word that is not in the lexicon or because it is not in the system's set of functionalities. A unimodal approach for providing different levels of feedback of understanding in situations with different levels of uncertainty at different processing levels has been suggested in [27]. For the context of smart homes such an approach would have to be extended towards using multiple modalities, e.g. provide visual feedback of the recognized words or commands, highlighting objects involved in the recognized command etc. (2) on the other hand anticipative transparency is needed which provides the user with information about possible commands s/he can use. The participants of this study suggested a—potentially contextualized—list of possible commands, either text-based or via verbal request. However, this leaves the potentially complex process of understanding what these commands actually mean to the user. This could be solved with a manual. Yet, manuals have the same problem: they often provide explanations that are outside the experience world of the users e.g. by using termini and explanations that are not adapted to the user's background. We therefore suggest research, to provide information that is capable of taking the user's perspective into account.

b) User's Perspective: We can divide these approaches into anticipative and interactive. (1) Anticipative approaches to transparency from the user's perspective entail the ability for the user to ask for a certain capability and the system being able to provide a feedback of commands that could potentially provide the requested functionality, e.g. based on latent semantic analysis or other semantic or text-based machine learning approaches. However, the most important capability would be for the system to provide negative feedback in case it realizes that a certain functionality can not be achieved, i.e. the system needs to know its own limitations. (2) Interactive approaches to transparency from the user's perspective entail the capability of the system to negotiate with the user which command s/he might need e.g. by providing exemplary executions of the commands. One goal here would be to teach the user to better understand the system's underlying ontology of functionalities. However, this still entails the user to adapt to the system. Therefore, (3) interactive learning approaches are needed that through joint learning enable the user to teach the system new functionalities, e.g. by demonstration, and the associated interactive means, e.g. by speech, gesture or specification of preconditions, to activate those. As learning significantly changes the system's behavior this approach requires specific interaction models. Also, these approaches can have two different targets: (i) process learning, where the user teaches specific sequences of actions that the system just associates with a certain command. In contrast, (ii) goal learning would entail a communication where not commands are given but the user specifies her or his goals, e.g. "I want to work". Such an approach provides the ability to extend already learned concepts, e.g. by observing user behavior after

s/he has announced to be working. This can entail the deactivation of communication channels when the cell phones rings or other reactions to changing contexts. Thus, learning can actually provide a path towards "Transparency by design" as the user directly experiences and influences the system's learning trajectory.

V. Discussion and Conclusion

Until the vision of smart homes as personalized and competent assistants for their indwellers becomes fulfilled, a number of problems and limitations have to be solved and overcome. In this article we presented a survey that takes stock of the current (as of late 2018) state-of-the-art interaction capabilities of PVAs using a qualitative evaluation of reported problems and wished features by users of today's PVA systems.

First of all, we need to acknowledge that the user group is highly skewed and far from allowing a generalization to the general population. The participants are dominantly male, tech-savy and belong to the WEIRD group ('western educated industrialized rich democratic') [28] of people, globally a small minority. Hence, conclusions have very limited potential for extrapolation. In addition, being (early) adopters of an upcoming technology, they may have little resistance against the risks and pitfalls and can basically be assumed to have a positive attitude towards such systems, otherwise they would not be using the systems nor be willing to participate in a study that promises to get another one of those systems (via the raffle). However, this would only be an issue if our interest had been into the balance between positive and negative issues. Since, however, we only focus on the users' issues (i.e. addressed problems and wished improvements), we can take the observations as a lower bound to assess the problems of PVAs, as average users will usually have more problems adapting to the systems—or knowledge deficits to use PVAs adequately. In awareness of these limitations, however, the study provides a quite substantial baseline, a reality check, of how current PVAs are capable of interaction. As all systems rely on language and language only, it is no surprise that language processing issues are the overwhelming number of identified issues. Particularly, as when using natural language, users can be expected to assume a matched cognitive competence of their interlocutor. Also, due to the lack of a camera, the inability to recognize body orientation, gestures, mimics or emotional display makes it harder for a system to associate the illocutionary act of a speech act.

Concerning authentication and authorization, many PVAs already have a voice recognition feature to authenticate the speaker during the conduction of our study, yet it was—and still is—not secure and it was not used for smart home control. Since our study in late 2018, Google introduced two-factor authentication into the smart home capabilities of the Google Assistant[5]. This allows users to set pin codes for controlling smart home appliances and is an important next step to fulfill part of the identified recommendations of our paper As to activity-based control, Google introduced "Custom Routines" for their Google Assistant [29] just one month before our study took place. So, at the time of our study this feature was still limited to the USA, which slightly affected our study results. *Routines* enable the execution of predefined commands via a customizable trigger sentence. Therefore, it is a further step towards activity-based control, since with routines, inhabitants can pre-configure the environment to common activities. Meanwhile, Google also uses the location (i.e. room) of their smart speakers to limit the scope of controlled devices to that room by default, adding to our recommendation of situatedness. Furthermore, Google worked on more natural interactions as well: they presented first approaches[6] for ongoing conversations, which should allow continued interactions in terms of follow-up requests. This is a first step towards more situated dialogue.

In summary, our investigation shows that even though smart homes and particularly PVAs become increasingly popular, there are still various issues and thus many opportunities for improvement. Based on the insights gained from our qualitative evaluation of current issues and requested improvements, we derived design recommendations for the future development of PVAs, along the dimensions *Authentication & Authorization*, *Activity-Based Interaction*, *Situated Dialogue*, and *Explainability & Transparency*. In our ongoing research on our 'Cognitive Service Robotics Apartment as Ambient Host' (CSRA), we continue to extend system components' capabilities with regard to these four guiding principles.

The collected data of our evaluation is published online at https://gitlab.ub.uni-bielefeld.de/thuxohl/pva-online-survey---results-and-evaluation.

Acknowledgment

This work was funded as part of the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277), Bielefeld University.

References

[1] R. Knote, A. Janson, M. Söllner, and J. M. Leimeister, "Classifying Smart Personal Assistants : An Empirical Cluster Analysis," in *HICSS Proceedings 2019*, 2019, pp. 2024–2033.

[2] P. Bahl, R. Y. Han, L. E. Li, and M. Satyanarayanan, "Advancing the state of mobile cloud computing," in *Proceedings of the third ACM workshop on Mobile cloud computing and services - MCS '12*. New York, New York, USA: ACM Press, 2012, pp. 21–28.

[3] Tractica, "The Virtual Digital Assistant Market Will Reach $15.8 Billion Worldwide by 2021," 2016, retrieved 2019-07-15. [Online]. Available: https://www.tractica.com/newsroom/press-releases/the-virtual-digital-assistant-market-will-reach-15-8-billion-worldwide-by-2021/

[4] S. Karpagavalli and E. Chandra, "A Review on Automatic Speech Recognition Architecture and Approaches," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 9, no. 4, pp. 393–404, 2016.

[5] W. Xiong, J. Droppo, X. Huang, F. Seide, M. L. Seltzer, A. Stolcke, D. Yu, and G. Zweig, "Toward Human Parity in Conversational Speech Recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 12, pp. 2410–2423, 2017.

[5]https://developers.google.com/actions/smarthome/two-factor-authentication, retrieved 2019-05-06

[6]https://www.blog.google/products/assistant/chatting-your-google-assistant-just-got-easier/, retrieved 2019-05-07

[6] M. Pohling, C. Leichsenring, and T. Hermann, "Base Cube One: A location-addressable service-oriented smart environment framework," *Journal of Ambient Intelligence and Smart Environments*, vol. 11, no. 5, pp. 373–401, sep 2019. [Online]. Available: https://ip.ios.semcs.net/articles/journal-of-ambient-intelligence-and-smart-environments/ais190533

[7] T. Eric, S. Ivanovic, S. Milivojsa, M. Matic, and N. Smiljkovic, "Voice control for smart home automation: Evaluation of approaches and possible architectures," *IEEE International Conference on Consumer Electronics - Berlin, ICCE-Berlin*, pp. 140–142, 2017.

[8] C. Olson and K. Kemery, "Voice Report 2019 - From answers to action: customer adoption of voice technology and digital assistants," 2019, retrieved 2019-07-15. [Online]. Available: https://about.ads.microsoft.com/en-us/insights/2019-voice-report

[9] R. Budiu and P. Laubheimer, "Intelligent Assistants Have Poor Usability: A User Study of Alexa, Google Assistant, and Siri," 2018, retrieved 2019-05-13. [Online]. Available: https://www.nngroup.com/articles/intelligent-assistant-usability/

[10] G. López, L. Quesada, and L. A. Guerrero, "Alexa vs. siri vs. cortana vs. google assistant: A comparison of speech-based natural user interfaces," in *Advances in Human Factors and Systems Interaction*. Cham: Springer International Publishing, 2018, pp. 241–250.

[11] A. Purington, J. G. Taft, S. Sannon, N. N. Bazarova, and S. H. Taylor, ""Alexa is my new BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo," in *CHI EA '17 Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2017, pp. 2853–2859.

[12] H. Williams, I. Knight, I. Lopatovska, K. Cosenza, K. Raines, K. Rink, A. Martinez, Q. Li, P. Sorsche, and D. Hirsch, "Talk to me: Exploring user interactions with the Amazon Alexa," *Journal of Librarianship and Information Science*, 2018.

[13] D. Caivano, D. Fogli, R. Lanzilotti, A. Piccinno, and F. Cassano, "Supporting end users to control their smart home: design implications from a literature review and an empirical investigation," *Journal of Systems and Software*, vol. 144, pp. 295–313, 2018.

[14] S. S. Lee, J. Lee, and K. P. Lee, "Designing Intelligent Assistant through User Participations," in *DIS '17 Proceedings of the 2017 Conference on Designing Interactive Systems*, 2017, pp. 173–177.

[15] A. Coskun, G. Kaner, and İdil Bostan, "Is Smart Home a Necessity or a Fantasy for the Mainstream User? A Study on Users' Expectations of Smart Household Appliances," *International Journal of Design*, vol. 12, no. 1, pp. 7–20, 2018.

[16] S. Wrede, C. Leichsenring, P. Holthaus, T. Hermann, and S. Wachsmuth, "The Cognitive Service Robotics Apartment," *KI - Künstliche Intelligenz*, vol. 31, no. 3, pp. 299–304, 8 2017.

[17] S. Team, "Personalized Hey Siri," *Apple Machine Learning Journal*, vol. 1, no. 9, 2018, retrieved 2019-05-13. [Online]. Available: https://machinelearning.apple.com/2018/04/16/personalized-hey-siri.html

[18] F. Portet, M. Vacher, C. Golanski, C. Roux, and B. Meillon, "Design and evaluation of a smart home voice interface for the elderly: Acceptability and objection aspects," *Personal and Ubiquitous Computing*, vol. 17, no. 1, pp. 127–144, 2013.

[19] M. B. Hoy, "Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants," *Medical Reference Services Quarterly*, vol. 37, no. 1, pp. 81–88, 2018.

[20] A. Tilley, "How A Few Words To Apple's Siri Unlocked A Man's Front Door," 2016, retrieved 2019-05-13. [Online]. Available: https://www.forbes.com/sites/aarontilley/2016/09/21/apple-homekit-siri-security/

[21] A. Liptak, "Amazon's Alexa started ordering people dollhouses after hearing its name on TV," 2017, retrieved 2019-05-13. [Online]. Available: https://www.theverge.com/2017/1/7/14200210/amazon-alexa-tech-news-anchor-order-dollhouse

[22] T. A. Nguyen and M. Aiello, "Energy intelligent buildings based on user activity: A survey," *Energy and Buildings*, vol. 56, pp. 244–257, 2013.

[23] Z. Prasov and J. Y. Chai, "What's in a Gaze? The Role of Eye-Gaze in Reference Resolution in Multimodal Conversational Interfaces," in *IUI '08 Proceedings of the 13th international conference on Intelligent user interfaces*, 2008, pp. 20–29.

[24] V. Richter, B. Carlmeyer, F. Lier, S. M. zu Borgsen, D. Schlangen, F. Kummert, S. Wachsmuth, and B. Wrede, "Are you talking to me?: Improving the robustness of dialogue systems in a multi party hri scenario by incorporating gaze direction and lip movement of attendees," in *HAI*, 2016.

[25] D. McColl, A. Hong, N. Hatakeyama, G. Nejat, and B. Benhabib, "A survey of autonomous human affect detection methods for social robots engaged in natural hri," *Journal of Intelligent & Robotic Systems*, vol. 82, no. 1, pp. 101–133, Apr 2016.

[26] B. Carlmeyer, D. Schlangen, and B. Wrede, "Exploring self-interruptions as a strategy for regaining the attention of distracted users," in *Proceedings of the 1st Workshop on Embodied Interaction with Smart Environments-EISE'16*, 2016.

[27] S. E. Brennan and E. A. Hulteen, "Interaction and feedback in a spoken language system: A theoretical framework," *Knowledge-based systems*, vol. 8, no. 2-3, pp. 143–151, 1995.

[28] J. Henrich, S. J. Heine, and A. Norenzayan, "Most people are not weird," *Nature*, vol. 466, no. 7302, p. 29, 2010.

[29] S. Huffman, "The future of the Google Assistant: Helping you get things done to give you time back," 2018, retrieved 2019-05-13. [Online]. Available: https://www.blog.google/products/assistant/io18/